

Even Flood, Bibliotekforum 2, 2006

Project Gutenberg

av Even Flood

Dette burde jeg egentlig ha skrevet om for lenge, lenge siden. Project Gutenberg, <http://www.gutenberg.org/> er et gammelt prosjekt. Det går tilbake til 1971 da Michael Hart, den gang student ved University of Illinois, fikk tilgang til datamaskinkapasitet verd 100 millioner dollar på en datamaskin og bestemte seg for å bruke den ressursen til noe verdifullt. Han så for seg en tid da maskiner ville bli knyttet til hverandre i et nettverk, Og begynte derfor å legge inn tekster av viktige dokumenter så de kunne bli tilgjengelige elektronisk. Han startet det han kalte for Project Gutenberg. Det første han la opp var den amerikanske uavhengighetserklæringen. Tilfeldigvis var denne maskinen knyttet til det nettverket som senere skulle bli til Internett. Resten er, som man sier, historie.

Project Gutenberg har vokst seg meget stort og er nå det største arkivet for elektroniske bøker på veven. Pr 1/1 2006 var det 18183 bøker, i tillegg er det også andre dokumenter. Nesten 2 millioner lastes ned hver måned. Målet er å legge inn 500 nye bøker i måneden. Arbeidet er basert på frivillige, spesielt for å få tak i bøker og til å korrekturlese dem. Bøkene skannes inn, men resultatet må alltid gjennom grundig korrektur.

Det er statistikk over bruken, og den kan være interessant: Mest nedlastede de siste månedene (pr 11/1/2006) var:

The Art of War (10941 nedlastinger)

The Notebooks of Leonardo Da Vinci — Complete by Leonardo da Vinci (10463)

Kamasutra (9549)

A Christmas Carol (7491)

Spesielt nr 2 på listen er overraskende, siden materialet bare er i tekstformat uten illustrasjoner. I begynnelsen var alt materialet lagret som tekstfiler i det mest primitive formatet, ascii. Det formatet er fremdeles det viktigste og standarden for alle dokumentene. Men det arbeidet med å omformatere det til mer standard vevformat, som XML og HTML. Mye mer om dette er på <http://gutenberg.hwg.org/>. En del bøker finnes også som lydfiler, de fleste lest med computerbasert tale. Det vil si at ordene er lest inn av et menneske og uttales ett for ett. Det er lagt opp for MP3 filer. Det er et meget godt tiltak. Men det er allikevel merkelig å høre "Tarzan" lest på den måten, den stemmen får selv de mest spennende bøker til å virke kjedelige.

Copyright har hele tiden vært et problem, naturlig nok. Bare dokumenter hvor de økonomiske rettighetene er gått ut blir lagt inn. Men det ble problemer da rettighetslovgivningen ble forandret underveis. I 1998 ble rettighetstiden forlenget fra 50 til 70 år etter opphavspersonens død. Unntak ble gjort for bøker som allerede var lagt inn i prosjektet. Det gjør at en del av forfatterinnførslene kan virke litt inkonsekvente siden ikke alt er lagt inn. Kvaliteten på tekstene, spesielt oversatte, kan også variere. Det er flere klassikere, som Jules Verne, som finnes i flere engelske oversettelser av meget varierende kvalitet og det er ikke sikkert den som legges inn er god. De aller fleste bøkene i Project Gutenberg er på engelsk, mange andre språk finnes også. Fransk, tysk og finsk(!) er de nasjonene som har flest verker etter engelsk. Bare åtte norske verker er lagt inn, seks av Ibsen og to av Wergeland.

Michael Hart er fremdeles meget aktiv i prosjektet med ukentlige nyhetsbrev. Et interessant intervju med ham er her: <http://samvak.tripod.com/busiweb46.html>. Det er også en god Wikipediaartikkel, http://en.wikipedia.org/wiki/Project_Gutenberg

Til slutt: Prosjekt Runeberg, <http://runeberg.org/>, er en nordisk avlegger av Project Gutenberg. Jeg lover å komme tilbake til det senere.